Remarks

By the present amendment, claims 1, 2, 4, 5, 7, 8, 10, 11, 13-16, 22 and 23 have been amended and no claims have been deleted, rendering 1-23 claims pending in the present application. The amendments to the claims have been made without prejudice and without acquiescing to any of the Examiner's objections. Applicant reserves the right to pursue any of the deleted subject matter in a further divisional, continuation or continuation-in-part application.

The Official Action dated November 20, 2002 has been carefully considered. It is believed that the amended claims submitted herewith and the following comments represent a complete response to the Examiner's rejections and place the present application in condition for allowance. Reconsideration is respectfully requested.

Election/Restriction Requirement

Claims 5, 11 and 13-15 have been amended to delete reference to SEQ ID NOS: 1, 4 and 6 in view of the Restriction Requirement.

Drawings

We are enclosing herewith revised Figures 1-4 and 9 to overcome the objection under 37 CFR §1.84(h)(2) that certain Figures are not labeled separately. Former Figures 1-4 and 9 have been relabeled Figures 1A-D; 2A-C; 3A-B; 4A-E and 9A-D.

Claim Objections - Informalities

The Examiner has requested that in Claims 5, 11, 14 (b)-(e) and Claim 15 (a) (1)-(4), the phrase "a nucleic acid sequence of" is replaced with "the nucleic acid sequence". While we believe it is not appropriate to amend the claims in every instance these claims recite "a nucleic acid sequence", we have amended these claims in accordance with the Examiner's request whenever appropriate. These claims have also been amended to ensure that they are not directed to non-elected inventions.

The Examiner has requested that in claims 14 and 15 the phrase "isolated nucleic acid sequence" is replaced with "isolated nucleic acid molecule". Claims 14 and 15 have been amended in accordance with the Examiner's request.

The Examiner has requested that in claims 22 and 23 the phrase "comprising a nucleic acid sequence" is replaced with the phrase "the nucleic acid sequence. Claims 22 and 23 have been amended in accordance with the Examiner's request.

Claims 5, 11, 14 and 15 have also been amended to correct the spelling of "complementary".

35 U.S.C. §112, first paragraph

(a) Written Description

The Examiner has objected to claims 1-23 under 35 U.S.C. §112, first paragraph as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. We respectfully disagree with the Examiner for the reasons that follow.

Applicants submit that the specification conveys with reasonably clarity to those skilled in the art that Applicants were in possession of the invention as now claimed at the time the application was filed. We will discuss the method and composition claims separately.

Composition claims 14-23 are limited to a flax promoter having a nucleic acid sequence shown in SEQ ID NO:8 (Figure 4). The claim also covers sequences that are complementary to the sequence as SEQ ID NO:8; sequences that have substantial sequence homology to SEQ ID NO:8; sequences that are analogs to SEQ ID NO:8 and sequences that hybridize to SEQ ID NO:8 under stringent hybridization conditions. Applicants were the first to isolate the promoter having the sequence shown in SEQ ID NO:8. As a result of Applicants' invention, one of skill in the art,

having read the disclosure of the present application, could readily isolate or prepare modifications to the sequence shown in SEQ ID NO:8 as provided in the claim. In particular, the disclosure provides on pages 10-12 examples of modifications that can be made to the sequence in order to prepare the claimed sequences. Further, we strongly submit that it would be unfair to limit the Applicants to the particular sequence as SEQ ID NO:8 as those skilled in the art could readily modify the sequence in order to circumvent the claim. In addition, Applicants have isolated four flax seed-specific promoters which is a representative number of species to demonstrate that Applicants are entitled to the scope of claim as currently pending. It is worth quoting from *in Re Goffe*:

> "to provide effective incentives, claims must adequately protect inventors. To demand that the first to disclose shall limit his claims to what he has found will work or to materials which meet the guidelines specified for "preferred" materials in a process such as the one herein involved would not serve the constitutional purpose of promoting progress in the useful arts." (*in Re Goffe*, 542 F.2d 564, 567, 191 USPQ 429, 431 (CCPA 1976)).

Claims 1-13 relate to methods for the expression of a nucleic acid sequence of interest in flax seeds using a seed-specific promoter obtained from flax as well as flax plants and flax seeds prepared by the method. We respectfully submit that the independent method claims 1, 7 and 13 do not need to be limited to particular flax promoters. As mentioned above, in the present application, the Applicants have illustrated the effectiveness of the method of their invention through using four different flax seed promoters. Accordingly, we submit that the description of four different promoters in the method of the invention is sufficient to indicate that Applicants have possession of the claimed invention.

We point out to the Examiner that in the training materials that were published on March 1, 2000 that accompanied the Written Description Guidelines, there is an example, Example 18, that addresses the situation wherein the invention relates to a

method that employs a nucleic acid molecule. In that case, they provided only one example with a specific nucleic acid molecule and it was held that "the single embodiment is representative of the genus". Consequently, in the present case, we respectfully submit that four embodiments are represented of the genus and that the claims meet the Written Description requirements.

(b) Enablement

The Examiner has objected to claims 1-23 under 35 U.S.C. §112, first paragraph, alleging that the specification is only enabling with respect to SEQ ID NO:8. We respectfully disagree with the Examiner for the reasons that follow.

The requirement of enabling disclosure does not mean that the applicant must describe all actual embodiments. How a teaching is set forth, by specific example or broad terminology, is not important (*in Re Marzocchi*, 439 F.2d 220, 223-24 169 USPQ 367, 370 (CCPA 1971)). As concerns the breadth of a claim relevant to enablement, the only relevant concern should be whether the scope of enablement provided to one skilled in the art by the disclosure is commensurate with the scope of protection sought by the claims (*in Re Moore*, 439 F.2d 1232, 1236, 169 USPQ 236, 239 (CCPA 1971)). The scope of enablement must only bear a "reasonable correlation" to the scope of the claims (*in Re Fisher*, 427F.2d 833, 839, 166USPQ 18, 24 (CCPA 1970)).

Claims 5, 11, 14-23 relate to SEQ.ID.NO.:8 and to nucleic acid sequences that (i) are hybridizing thereto under stringent conditions; or (ii) have substantial sequence homology therewith; or (iii) are complementary thereto; or (iv) are an analog thereof. The Examiner has objected that the application is not enabling as the claimed nucleic acid sequences are described solely in terms of their function and not their structure. Applicant disagrees with the Examiner and respectfully submits that there exists in the art to which the invention pertains a well recognized correlation between: (i) the similarity in chemical structure of nucleic acid molecules and the ability of nucleic acid molecules to hybridize under stringent conditions; and (ii) the similarity in chemical

structure of nucleic acid molecules and the degree of homology between nucleic acid molecules; and (iii) the similarity in chemical structure of nucleic acid molecules and the degree of complementarity between nucleic acid molecules; and (iv) the similarity in chemical structure of nucleic acid molecules and their analogs. To support this assertion Applicant herewith encloses the following textbook reference for the Examiner's consideration: Lewin B., 1994, Genes V Pages 111-113, which states *inter alia* that [*the ability of two nucleic acid sequences to hybridize constitutes a precise test for their complementarity since only* <u>*complementary sequences*</u> *can form a duplex structure…*] and [*….the complementarity between single strands can be used to indicate the similarity between the original duplex molecules*]. Furthermore Applicant points out that the terms "sequence that has substantial sequence homology", "sequence that hybridizes" and "a nucleic acid sequence which is an analog" have been defined in the specification to further clarify Applicant's intended meaning of these terms (see: Page 10, line 28 -Page 12 line 25). The claims recite nucleic acid molecules that (i) hybridize to SEQ.ID.NO.:8 under stringent hybridization conditions; or (ii) have substantial sequence similarity to SEQ.ID.NO.:8; or (iii) are complementary to SEQ.ID.NO.:8 or (iv) are an analog of SEQ.ID.NO.:8. The claims clearly do not recite <u>any</u> promoter capable directing seed-specific expression obtainable from flax.

The Examiner also points out that in certain instances limited nucleotide substitutions may result in significant functional changes (such as is the case Chamberland at al. and Donald et al. art cited by the Examiner). In response Applicant respectfully submits that such substitutions are the exception rather than the rule. It should be noted in this regard that the authors of the Chamberland et al. paper identified the legumin box within the promoter of the soybean β-conglycinin promoter as an element suspected to be important for promoter function prior to preparing promoter mutants comprising nucleotide substitutions <u>within</u> the legumin box and that furthermore the plant mutants that were obtained retain significant promoter activity. Similarly, in the Donald et al. paper mutations within the previously identified G, I, and GT boxes,

elements putatively important for promoter function within the *Arabidopsis thaliana* rbcS-1A promoter, were evaluated and again mutants typically retain promoter activity.

Thus we respectfully submit that the specification fully, clearly and concisely describes the claimed nucleic acid molecules and provides sufficient guidance to a person of ordinary skill in the art to make and use these molecules.

Claims 1-4, 6-10 and 12-13 are directed to <u>methods</u> for the expression of a nucleic acid sequence in of interest in flax seeds using a seed-specific promoter obtained from flax and the resultant flax plants and seeds. Applicant discloses 4 different seed-specific promoters isolated from flax. In addition the application teaches a person of ordinary skill in the art how to readily obtain additional seed specific promoters (see page 15, lines 10-32) and use such flax seed specific promoters in accordance with the present invention. Accordingly we respectfully submit that Applicant has demonstrated by using a representative number of seed specific promoters that such promoters are useful in the expression of a nucleic acid sequence under the control of a seed specific promoter in flax seeds. Applicant therefore is entitled to claim a <u>method</u> for the expression of a nucleic acid sequence using any flax seed specific promoter <u>in flax seeds</u>.

The Examiner has suggested to amend the claims to recite a "seed-preferred promoter" rather than a "seed-specific promoter". We thank the Examiner for his suggestion and have herewith amended the claims and the specification in accordance with this suggestion.

In view of the foregoing, we respectfully request that the objections to the claims under 35 U.S.C. §112, first paragraph be withdrawn.

<u>35 U.S.C. §112, second paragraph</u>

The Examiner has objected to claims 2-5, 8, 9, 11, 12 and 14-23 under 35 U.S.C. §112, second paragraph, as failing to particularly point out and distinctly claim the

subject matter which is regarded as the invention Our comments to these objections are as follows.

The Examiner has objected to claim 2 'because it is unclear to what the claim is referring as to "characteristic conferred by said seed-specific promoter" is "conferred to said non-native nucleic acid sequence"'. We respectfully disagree with the Examiner as the claim is clear in that it is the seed-specific promoter that is conferring the expression characteristic to the non-native nucleic acid sequence of interest as opposed to the native nucleic acid sequence conferring the characteristic to the non-native nucleic acid sequence. Claims 2 and 3 are meant to specify that the seed-specific promoter confers a characteristic that it would normally confer on its native sequence to the nucleic acid sequence of interest.

The Examiner has objected to claim 4 for being improper Markush format. In response Applicant has herewith amended the claims to recite "is selected from the group of promoters consisting of..."

The Examiner has objected to claims 5, 11, 14 and 15 reciting "has substantial sequence homology to; "is an analog of a nucleic acid sequence" and "hybridizes under stringent hybridization conditions". Applicant agrees with the Examiner that all of these terms potentially could be unclear to a person of skill in the art, however we respectfully submit that in conjunction with the precise definition of each of these terms as set forth in the specification from Page 10, line 28 - Page 12, line 25 these terms will be readily understood by the skilled artisan. Thus the claims particularly point out and distinctly claim the subject matter which the applicant regards as his invention.

The Examiner has objected to claim 15 as being indefinite as it is unclear how the nucleic acid sequence at 15(a)(2) could hybridize to the nucleic acid sequence of 15(a)(2), itself. We agree with the Examiner and have revised the claims in

accordance with the Examiner's suggestion so that 15(a)(2) recites 15(a)(1); 15(a)(3) recites 15 (a)(1) or 15 (a)(2); and 15(a)(4) recites 15(a)(1) or 15(a)(2) or 15(a)(3).

In view of the foregoing, we respectfully request that the objections to the claims under 35 U.S.C. §112, second paragraph be withdrawn.

<u>35 U.S.C. §102</u>

The Examiner has objected to Claims 1,3, 5-9 and 11-23 as being anticipated by Jain et. (WO 98/18948). We respectfully disagree with the Examiner for the reasons that follow.

Jain et al. discloses two flax promoters sequences operably linked to two stearoyl-acyl carrier protein desaturase (SAD) coding sequences from flax. In order for Jain et al. to anticipate the invention the disclosure must provide each and every element of the claim. While the SAD promoters disclosed in Jain et al. are capable of directing the expression of heterologous nucleic acid sequences in seed, significant expression is observed in other tissues as well. For example, Fig. 6 shows significant expression of SAD2 in young leaves and apices; mature leaves; stems; buds; half open flowers and flowers and Fig. 10 shows GUS activity, which, according to Jain et al. could be [...*easily detected in both leaves and seeds*] (see Page 21, Line 21). Furthermore Jain et al. state that *[...these promoters are useful in manipulating transgene expression in variety of tissues including seed*] (see: page 9, lines 25-26) and [...*these promoters were capable of expressing the uidA gene in various tissues*...] (see: page 21, line 6-7). Thus a person of skill in the art would when following the teachings of Jain et al. not expect to achieve expression seed-specific or seed preferred expression, as such term will be understood by a person of skill in the art having read the instant specification. In *Scripps Clinic & Research Foundation v. Genentech, Inc.* (927 F.2d 1565, 18 USPQ 2d 1001 Fed Cir:1991), the Court held that in order for there to be anticipation, the prior art must place the invention in the possession of the public by providing an enabling disclosure of how to make and use the claimed subject matter. Jain et al. clearly does not enable the production of <u>seed-</u>

specific expression of a nucleic acid sequence of interest. Consequently, Jain et al. cannot be said to anticipate the claims of the invention.

In view of the foregoing, we respectfully request that the objections to the claims under 35 U.S.C. §102 be withdrawn.

Attached hereto is a marked-up version of the changes made to the specification and claims by the current amendment. The attached page is captioned "Version with markings to show changes made".

The Commissioner is hereby authorized to charge any fee (including any claim fee) which may be required to our Deposit Account No. 02-2095.

In view of the foregoing comments, we respectfully submit that the application is in order for allowance and early indication of that effect is respectfully requested. Should the Examiner deem it beneficial to discuss the application in greater detail, he is kindly requested to contact the undersigned by telephone at (416) 957-1682 at his convenience.

Respectfully submitted,

Sarita. Chaudhary et al.

M. Gravelle

Micheline Gravelle
Registration No. 40,261

Bereskin & Parr
Box 401, 40 King Street West
Toronto, Ontario, Canada M5H 3Y2

(416) 364-7311

Version with markings to show changes made

In the Specification

Page 6, lines 24-32 have been amended as follows:

--Figures 1A-D show[s] the DNA sequence (SEQ.ID.NO.:1) of a flax genomic clone encoding a 16.0 kDa oleosin protein (SEQ.ID.NOS.:2 and 3).

Figures 2A-C show[s] the DNA sequence (SEQ.ID.NO.:4) of a flax genomic clone encoding a 18.6 kDa oleosin protein (SEQ.ID.NO.:5).

Figures 3A-B show[s] the DNA sequence (SEQ.ID.NO.:6) of a flax genomic clone encoding a 2S storage protein (SEQ.ID.NO.:7).

Figures 4A-E show[s] the DNA sequence (SEQ.ID.NO.:8) of a flax genomic clone encoding a 54.5 kDa legumin-like storage protein (SEQ.ID.NOS.:9-12).--

Page 7, lines 9-11 have been amended as follows:

--Figures 9A-D show[s] GUS expression in developing flax embryos and Arabidopsis seeds of plants transformed with a 2S protein gene promoter GUS fusion.--

Page 10, lines 7-10 have been amended as follows:

-- The terms "seed-specific promoter" or "seed-preferred promoter", both of which terms may be used interchangeably herein, mean that a gene expressed under the control of the promoter is predominantly expressed in plant seeds with no or no substantial expression, typically less than 5% of the overall expression level, in other plant tissues --.

In the Claims

Claims 1, 2, 4, 5, 7, 8, 10, 11, 13-16, 22 and 23 have been amended as follows:

1. *(amended)* A method for the expression of a nucleic acid sequence of interest in flax seeds comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

(1) a seed-preferred [specific] promoter obtained from flax; and

(2) said nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said seed-preferred [specific] promoter;

(b) introducing said chimeric nucleic acid construct into a flax plant cell; and

(c) growing said flax plant cell into a mature flax plant capable of setting seed

wherein said nucleic acid sequence of interest is expressed in the seed under the control of said seed-preferred [specific] promoter.

2 (*amended*). The method according to claim 1 wherein at least one expression characteristic conferred by said seed-preferred [specific] promoter to its native nucleic acid sequence is conferred to said non-native nucleic acid sequence.

4. (*amended*) The method according to claim 1 wherein said flax seed-preferred [specific] promoter is selected from the group of promoters consisting of [comprising,] oleosin promoters, 2S storage protein promoters and legumin-like seed-storage protein promoters.

5. (*amended*) The method according to claim 1 wherein said flax seed-preferred [specific] promoter comprises:

(a) the [a] nucleic acid sequence as shown in [Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or] Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

(b) a nucleic acid sequence that is complementary [complimentary] to the nucleic acid sequence of (a);

(c) a nucleic acid sequence that has substantial sequence homology to the nucleic acid sequence of (a) or (b);

(d) a nucleic acid sequence that is an analog of the nucleic acid sequence of (a), (b) or (c); or

(e) a nucleic acid sequence that hybridizes to the [a] nucleic acid sequence of (a), (b), (c) or (d) under stringent hybridization conditions.

7. (*amended*) Transgenic flax seed prepared according to a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

 (1) a seed-preferred promoter obtained from flax; and

 (2) a nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said seed-preferred promoter;

(b) introducing said chimeric nucleic acid construct into a flax plant cell; and

(c) growing said flax plant cell into a mature flax plant capable of setting seed

wherein said nucleic acid sequence of interest is expressed in the seed under the control of said seed-preferred [specific] promoter.

8. (*amended*) Transgenic flax seed according to claim 7 wherein at least one expression characteristic conferred by said seed-preferred [specific] promoter to its native nucleic acid sequence is conferred to said non-native nucleic acid sequence.

10. (*amended*) Transgenic flax seed according to claim 8 wherein said seed-preferred [specific] promoter is a seed storage protein promoter, an oleosin promoter, a 2S storage protein promoter or a legumin-like seed-storage protein promoter.

11. (*amended*) Transgenic flax seed according to claim 8 wherein said seed-preferred promoter comprises:

(a) the nucleic acid sequence as shown in [Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or] Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

(b) a nucleic acid sequence that is complementary [complimentary] to the nucleic acid sequence of (a);

(c) a nucleic acid sequence that has substantial sequence homology to the nucleic acid sequence of (a) or (b);

(d) a nucleic acid sequence that is an analog of the nucleic acid sequence of (a), (b) or (c); or

(e) a nucleic acid sequence that hybridizes to the [a] nucleic acid sequence of (a), (b), (c) or (d) under stringent hybridization conditions.

13. (*amended*) Transgenic flax plants capable of setting seed prepared by a method a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

(1) a seed-preferred [specific] promoter obtained from flax; and

(2) a nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said seed-preferred [specific] promoter;

(b) introducing said chimeric nucleic acid construct into a flax plant cell; and

(c) growing said flax plant cell into a mature flax plant capable of setting seed

wherein said nucleic acid sequence of interest is expressed in the seed under the control of said seed-preferred [specific] promoter.

14. (*amended*) An isolated nucleic acid molecule capable of directing seed-preferred [specific] expression in a plant comprising:

(a) the [a] nucleic acid sequence as shown in [Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or] Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

(b) the nucleic acid sequence that is complementary [complimentary] to the nucleic acid sequence of (a);

(c) a nucleic acid sequence that has substantial sequence homology to the nucleic acid sequence of (a) or (b); or

(d) a nucleic acid sequence that is an analog of the nucleic acid sequence of (a), (b) or (c); or

(e) a nucleic acid sequence that hybridizes to the [a] nucleic acid sequence of (a), (b), (c) or (d) under stringent hybridization conditions.

15. (*amended*) An isolated chimeric nucleic acid molecule comprising:

(a) a first nucleic acid sequence comprising a seed-preferred promoter obtained from flax which comprises:

(1) the nucleic acid sequence as shown in [Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or] Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

(2) a nucleic acid sequence that hybridizes to the nucleic acid sequence of (a)(1) under stringent hybridization conditions;

(3) a nucleic acid sequence that is complementary [complimentary] to the nucleic acid sequence of (a)(1) or (a)(2); or

(4) a nucleic acid sequence that has substantial sequence homology to the nucleic acid sequence of (a)(1); (a)(2) or (a)(3); and

(b) a second nucleic acid sequence non-native to said flax seed-preferred [specific] promoter.

16. (*amended*) A method for the expression of a nucleic acid sequence of interest in a plant seed comprising:

(a) introducing the chimeric nucleic acid molecule according to claim 15 into a plant cell; and

(b) growing said plant cell into a mature plant capable of setting seed,

wherein the second nucleic acid sequence is expressed in the seed under the control of the seed-preferred [specific] promoter.

22 (*amended*). A recombinant expression vector comprising the [a] nucleic acid sequence according to claim 14.

23. (*amended*) A recombinant expression vector comprising the [a] nucleic acid sequence according to claim 15.

FIGURE 1A

```
   1 ttcaaaacccgattcccgaggcggccctattgaagatatggggaagttcgacgagatcgatgtcgggtcgagtgctatg 80

  81 gtgatggtgccgtttgggggaggatgagcgagatagccaagactagcattccgttcccacacagagttgggaatttgta 160

 161 ccaaatccaacacttgtcgtattggagcgacgatagggacgcggaaaaacacatccgttggatcagggagttgtacgatg 240
                            R1
 241 atctcgagccttatgtgtcgaagaatccgaggtatgcttacgtgaactacagggatctcgacatcgggatgaatggagga 320

 321 ggtgaaggggatgagaagggtacttatggtgaggctaaggtgtgggggagaagtactttggggtcaactttgatcggtt 400
                                                                              R2
 401 ggttcgggtgaagacgattgttgatcccaataatgtgtttcgaaacgagcagagcattccctcaattccaactcggttat 480

 481 aaggatcaatgatcaatgagaattttcctttccaatgtgattacaagttctattgggtcagctttctcaactgctcctat 560

 561 tcatttagattaattcataacaactattaatttaccagccttttatccggcccgttggccgatttattttcttaagtttt 640

 641 agatgaaatgaaaccgatttagtttttattgagatgagattaatcttaatttgcttgaaatttactcacggttgatgtga 720
                                                 R3
 721 tatttggaattaactaaaatgataaatatcggataaaaataaaaatatttaaaataaataacataaacataagaacaata 800
         I1              I1                                 R4
 801 aaataaataaatttaattttaatttatttccttgtttctttctgtatcatacatctcttctcttacttcttaaaggctt 880
       R4
 881 ttcaattatcacttaattaaatacaatagataaatcgttaattctataacattaacctatacacttgcacggtgaacaat 960
                                                                              R5
 961 caatatgataatataataataatataataattcaattattaatctacaatttttaattataaagtttatgcggtcagtt 1040

1041 tctgcaagctccgagctccttgtcatcgttagtttctgcggtctcaaggtataacgactcggagcgacgagccctttgct 1120
                                                                   R1
1121 tccaatggacgggttgcatttctgccgtcgttgagctcgattggcgtgtcatgctggagtcagagttcctacaaaaaaac 1200

1201 cctaaactagagggtgattagggtgaaattagggtgttggcctgggttccattgtccaaagttttagtcaacttaaaaac 1280

1281 agacttaaattttatgcttcaaaatagtttatctgttattatattagcgtgtaattagtcttgacaatggggccggacgg 1360
```

FIGURE 1B

```
1361 gtacggattcgggaccccgatccccgcccatagtgtaatggctcaactgccaagtcagcattggaccgaaattattggac 1440

1441 acgaagtactaatgtgaaaactttacatttgttatttctactttaatactatgctattttcaaatttgaactttaat 1520
                                                          R6                    R6
1521 actatgttttatatagtttagtatatcttaattttatgcaaattcatctaattgtattaaactattttcgatccgtag 1600
                                R3
1601 ctaattatttcgaaggcaagtcaaagtgttattgtggactatgtgagctaatattgaacctttatctctcccaaccactc 1680

1681 aagttaattgaaccaaactcgatcggttgggtttcgagctatttcgagccattgttgttatatgcacgtgagatatcaag 1760
                            R2
1761 attgacccgaacactttattatgataatgtagaaaaagaaaacatattctaagactacatgcatgcaaagtgcaacccct 1840
                             I2                  I2
1841 gcatggaaagctgctcaacacgtggcatagactcccgccacgtgtccattccacctcatcacctcacccccaccgttcac 1920
                    CACA     ABRE
1921 ctcttattatatcacaacaatcaatcaatcctactcctccatactcgaacaaatccgaccaacttataccaatattccca 2000
         TATA         R5                                                   CAT
2001 aacttgattaatttctcagcaat ATG GAT CAG ACG CAC CAG ACA TAC GCC GGA ACC ACG CAG AAC  2065
   1                        M   D   Q   T   H   Q   T   Y   A   G   T   T   Q   N    14

2066 CCG AGC TAT GGC GGC GGG GGC ACA ATG TAC CAG CAG CAG CAG CCG AGG TCT TAC CAG GCG  2125
  15 P   S   Y   G   G   G   G   T   M   Y   Q   Q   Q   Q   P   R   S   Y   Q   A    34

2126 GTG AAG GCG GCC ACT GCA GCC ACC GCG GGT GGA TCC CTC ATC GTT CTG TCC GGT CTC ATC  2185
  35 V   K   A   A   T   A   A   T   A   G   G   S   L   I   V   L   S   G   L   I    54

2186 CTT ACG GCC ACC GTC ATT TCA CTC ATC ATA GCC ACC CCT CTC CTT GTC ATC TTC AGC CCT  2245
  55 L   T   A   T   V   I   S   L   I   I   A   T   P   L   L   V   I   F   S   P    74

2246 GTT CTT GTC CCG GCT CTC ATC ACC GTC GGG CTC TTG ATC ACC GGG TTT CTT GCT TCC GGT  2305
  75 V   L   V   P   A   L   I   T   V   G   L   L   I   T   G   F   L   A   S   G    94

2306 GGG TTC GGA GTC GCC GCC GTC ACC GTC TTG TCC TGG ATC TAT AG gtatgtataagctttggactt 2370
  95 G   F   G   V   A   A   V   T   V   L   S   W   I   Y   R                       109

2371 tagtattgttataaaatacataagctgatttatgaacatggatctcccaacaagagttatttaaatgcattctcggtctg 2450
```

FIGURE 1C

```
2451 actcgatcggttgggttttgagctactcggtcacaatggtcgggtcggctctggatctgttatactaatatttggaagcc 2530

2531 tgaagtttcattgttctgccccaacttcccactaccttttgagggtgttaagaagccatacaaactaattatgaatccct 2610

2611 cccaacaactcagaactcgagtcagtgggttgtgacggttctctataaacatttcgaaaatctttgttcaatgaacgtag 2690

2691 aaatgaccatgcttgatgattgtgggtcttataag G TAC GTG ACC GGC GGG CAC CCG GCG GGA GGG    2756
 110                                     Y   V   T   G   G   H   P   A   G   G      119

2757 GAT TCG CTG GAC CAG GCT AGG TCG AAG CTG GCC GGA AAG GCC AGG GAG GTG AAG GAC AGG 2816
 120 D   S   L   D   Q   A   R   S   K   L   A   G   K   A   R   E   V   K   D   R   139

2817 GCG TCG GAG TTC GCA CAG CAG CAT GTC ACA GGT GGT CAA CAG ACC TCT TAA agagagtcctct 2879
 140 A   S   E   F   A   Q   Q   H   V   T   G   G   Q   Q   T   S   *               156

2880 agttaaattggtcttcgtttctgtttcgtggcggcttgtaaactctcttttaagtgtgctgttttccttttgtctcgtgt 2959

2960 gttgtaagtgaaagtgtaatcgaagttccaagttggagatgtttgtaacgatgatgtttctaataatcagagatattaa 3039
                                                                Poly A signal
3040 aagggttgctaatttagtattgcgtctgatctcggaccaaactcgcaagtaaaattgcagaggatgagttgtacagaaca 3119

3120 agcgtgcattgttctggaagttcatctccttggagccgaccttgttgcttgcagtttcgccaagtccactagacaatgtt 3199

3200 acgagttaagcctctgtcaaacagatcgctctagcgtcccagaaaacaccagattttcgaaaaccatcggggatcaatt 3279

3280 ttcgattcaattccgatcttggaagtacttgaacagaagcatgatgctaaaagataatagaaaatcgaagcctagaaaag 3359

3360 ttgtacagaaagcaacaagtcaaaaatatagatcaacttcaaaggttcaaattacatcttacagaccccaaaaaatgaca 3439

3440 gttaacagaagtcgactaaacagaaaccagccagcttcacctggaatgaaggagctttgatcaatccatcctagcttcat 3519

3520 tcccctttgaaattgcagacagagctctcatcctgctaaagctggtggcttattcttaaccctgcaatcaataagcatga 3599

3600 actaacattggacaccttcatcggcggattgctcgaaaatcagtgagcgagggatttacctgtgtgtgtagtaacctctc 3679
```

FIGURE 1D

3680 tccttgtacataaaatctggaaattccggcatcaactactgccacctttctgcttaaggtgattttatcaccaaggctga 3759

3760 gcgtgattccttgcgtcttgctccgaatcctgatgtatccactgagctttccatctccttccttctccaggcttatgttc 3839

3840 accaatgcgtcctcgccgaacacactcttggcgtacaagttcgcagccaggaatccacactctccatcaagtgcagacct 3919

3920 gcaaaccccaaataagaacacaaactccaaagtcaacgatcaattctccgccttttatgaagaaaaggaaacttctgggt 3999

4000 acttacggtgccgtcagacacttcatatttgtagacttgatgatatggtccaggaattccttctcgttctgaattgttgt 4079

4080 gttaacagcaacctgacagacagaaagatatcgcaaatttaagatactgggatgactaggcacagagaaatgaaatctaa 4159

4160 ttctagaagtaaaaccttatttcccattcaaattctgcccacatagtccggaacgcagcatccgagcaagaagcaggag 4239

4240 agatgtaatccatgatatcgatgtggatatcgttgaggacgacaactgaacgttccatcacattgg 4305

FIGURE 2A

```
1    tctagacatttgacataaaccgaattcaaagaacacaacattgactaacaccaaaaagaaatagagtagtgaaatttgga 80
                  R1
81   agattaaaaaatagaaacaaactgattcttagaaagaagagatgattaggtgctttcagttcggtctgtcaggaaatcga 160
     R2
161  gatgttcacttatttacattgtcgattcatctcccaattgtcctggttcctttactgtccgacgctttttttgaatcccag 240
                       R3
241  ttaattcccatcaagtcttccttcagctgcgtagcactgctagctccaacatggagcgtggagtctactcgttcatgggg 320
321  catcgcaaaggtttgccttcatgttctgctaccagccagcgcccaccgcctcttggttgtgtggacaattgcggtgaagc 400
401  gcgcaagttgacatcccatagtctcgacacttcaccatatggatgtttaaaacgtatatcacgagtgcgatctacatgtc 480
481  ccatcacaccacatataaagcaatagtttgggagcttttcatatttgaaacgggcattgacgacttgccctctcgataat 560
561  ttaatctttttttctcttcagctgattgtgtgcatccattcgggctcagaagcacatcaaagggatctctccatcgtagt 640
641  attgggtcgtgtcgtatgatacgaagcagtcgatgaagtttcctaatgtgcgagctacaggctccgcaaagaacccgcga 720
721  ggtagatcgtatgctagtacccaaaaatcagtttgtcgtagcggaatcaacactagagactcaccctaatgcatctcatg 800
801  tgtgatgaacagtttatcatttgtgagtctaggggtcattgtcgatgacccaatgcacattgagcttatgatagaatttg 880
                                            R3
881  aataggaagcgttttccacccagatcacgaatagctacccctttttcgggcgccaaatttccggcatcctatcttccacc 960
961  acaacttaaagatgcgatcggtaaggaactcaccgaccacacacatcgaataatcttcggtgaccggttcctgttgatca 1040
1041 agtccctcaatttcctcaacctagtcttcaatcgccgctagcgttatcccccgcatatggactttcatagcgcggagcgt 1120
1121 agccggagacgacgagcaagaaggatgagcggcggcagattgcggctaaagaaacgagcttcctgccttgctctatggag 1200
1201 gcagatttctgagttgatggtgatggatttgtgatgtggacactttaatttaagttgatttttagcacttcattcacg 1280
1281 taattaaataaataatttccagtattttatatttatttccttacgttatctaattttttgaaagattaaaactttgatat 1360
     R4                                                         R2
```

FIGURE 2B

```
1361 aggcaagatcatgacacgtcgaagttaagtgaatgagactcctaacaaggtaataacaaagcagttcataaaccgaatga 1440
                                                                              R1
1441 ccttgatctttactaagcttgagatcattgaacatataattaaatacgttaatgaaagataagaactttaatataaaaat 1520
                                             R4
1521 cattcaaaacgagaaactgataacaaaaacaaagcaaacggccaacaaaataatagacggtggaaggatgatgcagagcc 1600
                                                                                R5
1601 atccacccttttttcccagtttccttactgcttacttctctatgcatatcacaagacgcccttgaaacttgttagtcatg 1680

1681 cagagcccttactcgccaggtcaccgcaccacgtgttactctatcacttctcctccctttcctttaaagaaccaccacgc 1760
       R5
1761 cacctccctctcacaaacactcataaaaaaaccacctcttgcatttctcccaagttcaaattagttcacagctaagcaag 1840

1841 aactcaacaaca ATG GCG GAT CGT ACA ACA CAG CCA CAC CAA GTC CAG GTC CAC ACC CAG CAC 1903
   1             M   A   D   R   T   T   Q   P   H   Q   V   Q   V   H   T   Q   H   17

1904 CAC TAT CCC ACC GGC GGG GCT TTC GGC CGT TAT GAA GGT GGA CTC AAA GGC GGT CCA CAT  1963
  18 H   Y   P   T   G   G   A   F   G   R   Y   E   G   G   L   K   G   G   P   H    37

1964 CAC CAG CAA GGA TCA GGC AGC GGC CCA TCA GCT TCC AAG GTG TTA GCA GTC ATG ACC GCG  2023
  38 H   Q   Q   G   S   G   S   G   P   S   A   S   K   V   L   A   V   M   T   A    57

2024 CTC CCC ATC GGC GGG ACC CTC CTT GCC TTG GCC GGG ATA ACC TTG GCT GGG ACG ATG ATC  2083
  58 L   P   I   G   G   T   L   L   A   L   A   G   I   T   L   A   G   T   M   I    77

2084 GGG CTG GCG ATC ACC ACC CCG ATT TTT GTC ATC TGC AGC CCT GTT CTA GTC CCG GCC GCT  2143
  78 G   L   A   I   T   T   P   I   F   V   I   C   S   P   V   L   V   P   A   A    97

2144 CTG CTC ATC GGG TTT GCC GTG AGC GCG TTT CTG GCC TCG GGG ATG GCC GGG CTG ACA GGG  2203
  98 L   L   I   G   F   A   V   S   A   F   L   A   S   G   M   A   G   L   T   G    117

2204 CTG ACC TCG CTG TCG TGG TTT GCG AGG TAT CTG CAG CAG GCT GGG CAG GGA GTT GGA GTG  2263
 118 L   T   S   L   S   W   F   A   R   Y   L   Q   Q   A   G   Q   G   V   G   V    137

2264 GGG GTG CCG GAT AGT TTC GAG CAG GCG AAG AGG CGC ATG CAG GAT GCT GCT GGG TAT ATG  2323
 138 G   V   P   D   S   F   E   Q   A   K   R   R   M   Q   D   A   A   G   Y   M    157
```

FIGURE 2C

```
2324 GGG CAG AAG ACC AAG GAA GTT GGG CAG GAG ATC CAG AGG AAG TCT CAG GAT GTG AAA GCA  2383
 158 G   Q   K   T   K   E   V   G   Q   E   I   Q   R   K   S   Q   D   V   K   A    177

2384 TCA GAC AAA TAA ggtgataataaggggttttgggttcgtgtgtaaactggtaaaatggaaattctgggttttactg 2459
 178 S   D   K   *                                                                     181

2460 tacttttgcatgtagtggaatgaatgagttcttgttctcttttgtcttttaatcataaagtaagaagcagcatttcatgt 2539

2540 tctggttgaatattgtcaagaattcgcaacaaatttagctaaaccagttcaatcttaccggttagacgacttcccagtaa 2619

2620 gaaacattccaggtccatcccggtataagagtctggacttctgaaacctttagaccttggatttggaaaaaagatgaaac 2699

2700 ctttagaataaattacaacgatggcagattgtacaaaactggagtcgagatcatgtaaattagcccataactaagaaccg 2779

2780 gcgatgacaacaattactaggaatatggttgttgggctggtcggcggctagcggtgatgatttggaagaatcggggatcc 2859

2860 agaatgtgagaaccgaatcatcgacgaacattacccggcgaggagcccatttcaagcaactttggaactcctatatggct 2939

2940 gttccagcaggccacctgctcaagaaagaaagaagccatgtcagaaatccttacgaaatctaactggatgctgatatgaa 3019

3020 tccgccaggtgtgcggagttctttacaggcaggatctataaagaagaaacatgttttgtattggcattgttgatgttcca 3099

3100 agcacgcagcgatctatctccggatcctaacaacaaaaatacggattctgtaagaaacaagcgcagaaaacttctgcaac 3179

3180 gaaaccactcgtatatttggttctgagttggagaaagatgaccatactactgtatttggttgaacttggattggaaccga 3259

3260 aattttgagttgaaaagcgagtgatcgtatataaatttcagattcagattaggatatcctatgagagaaggtagagttac 3339

3340 ctgatactacatactgcccatcaggggtaaaagttgcctcgatggttgtgtttggagatggttccaggctaaatccacaa 3419

3420 cgctgaacaaattaaaagatgaatggatcaatcttcaacccttacttctgcatttatgaggattggctcaaggctctcta 3499

3500 ga                                                                                3501
```

FIGURE 3A

```
  1  tccactatgtaggtcatatccatcattttaattttgggcaccattcaattccatcttgcctttagggatgtgaatatga   80
        5' primer (1)         AT rich
 81  acggccaaggtaagagaataaaaataatccaaattaaagcaagagaggccaagtaagataatccaaatgtacacttgtca  160
                         AT rich
161  tcgccgaaattagtaaaatacgcggcatattgtattcccacacattattaaaataccgtatatgtattggctgcatttgc  240

241  atgaataatactacgtgtaagcccaaaagaacccacgtgtagcccatgcaaagttaacactcacgaccccattcctcagt  320
       RY                                  G box  seed-specific
321  ctccactatataaacccaccatccccaatcttaccaaacccaccacacgactcacaactcgactctcacaccttaaagaa  400
           TATA                                                    3' primer (1)
401  ccaatcaccaccaaaaaATGGCAAAGCTGATGAGCCTAGCAGCCGTAGCAACGCAGTTCCTCTTCCTGATCGTGGTGGAC  480
  1                    M  A  K  L  M  S  L  A  A  V  A  T  Q  F  L  F  L  I  V  V  D    21

481  GCATCCGTCCGAACCACAGTGATTATCGACGAGGAGACCAACCAAGGCCGCGGTGGAGGCAAGGTGGCAGGGACAGCAGC  560
 22  A  S  V  R  T  T  V  I  I  D  E  E  T  N  Q  G  R  G  G  G  K  V  A  G  T  A  A    48

561  AGTCTGCGAGCAGCAGATCCAGCAGCGAGACTTCCTGAGGAGCTGCCAGCAGTTCATGTGGGAGAAAGTCCAGAGGGGCG  640
 49   V  C  E  Q  Q  I  Q  Q  R  D  F  L  R  S  C  Q  Q  F  M  W  E  K  V  Q  R  G  G   75

641  GCCACAGCCACTATTACAACCAGGGCCGTGGAGGAGGCGAACAGAGCCAGTACTTCGAACAGCTGTTTGTGACGACCTTA  720
 76   H  S  H  Y  Y  N  Q  G  R  G  G  G  E  Q  S  Q  Y  F  E  Q  L  F  V  T  T  L    101

721  AGCAATTGCGCACCGCGGTGCACCATGCCAGGGGACTTGAAGCGTGCCATCGGCCAAATGAGGCAGGAAATCCAGCAGCA  800
102  S  N  C  A  P  R  C  T  M  P  G  D  L  K  R  A  I  G  Q  M  R  Q  E  I  Q  Q  Q   128

801  GGGACAGCAGCAGGGACAGCAGCAGGAAGTTCAGAGGTGGATCCAGCAAGCTAAACAAATCGCTAAGGACCTCCCCGGAC  880
129   G  Q  Q  Q  G  Q  Q  Q  E  V  Q  R  W  I  Q  Q  A  K  Q  I  A  K  D  L  P  G  Q   155
```

FIGURE 3B

```
881  AGTGCCGCACCCAGCCTAGCCAATGCCAGTTCCAGGGCCAGCAGCAATCTGCATGGTTTTGAaggggtgatcgattatga 960
156   C  R  T  Q  P  S  Q  C  Q  F  Q  G  Q  Q  Q  S  A  W  F  *   5'primer (2)      175

961  gatcgtacaaagacactgctaggtgttaaggatggataataataataatgagatgaatgtgttttaagttagtgtaa 1040

1041 cagctgtaataaagagagagagagagagagagagagagagagagagagagagagagagaggctgatgaaatgttat 1120

1121 gtatgtttcttggtttttaaaataaatgaaagcacatgctcgtgtggttctatcgaattattcggcggttcctgtgggaa 1200

1201 aaagtccagaagggcggccgcagctactactacaaccaaggccgtggaggagggcaacagagccagcacttcgatagctg 1280

1281 ctgcgatgatcttaagcaattgaggagcgagtgcacatgcaggggactggagcgtgcaatcggccagatgaggcaggaca 1360

1361 tccagcagcagggacagcagcaggaagttgagaggtggtcccatcaatctaaacaagtcgctagggaccttccgggacag 1440

1441 tgcggcacccagcctagccgatgccagctccaggggcagcagcagtctgcatggttttgaagtggtgatcgatgagatcg 1520

1521 tataaagacactgctaggtgttaaggatgggataataagatgtgttttaagtcattaaccgtaataaaaagagagagagg 1600

1601 ctgatggaatgttatgtatgtatgtttcttggtttttaaaattaaatggaaagcacatgctcgtgtgggttctatc    1676
                                                             3'primer (2)
```

FIGURE 4A

```
        10        20        30        40        50        60        70        80        90       100
ctcaagcatacggacaagggtaaataacatagtcaccagaacataataaacaaaaagtgcagaagcaagataaaaaaattagctatggacattcaggttc

       110       120       130       140       150       160       170       180       190       200
atattggaaacatcattatcctagtcttgtgaccatccttcctcctgctctagttgagaggccttgggactaacgagaggtcagttgggatagcagatcc

       210       220       230       240       250       260       270       280       290       300
ttatcctggactagcctttctggtgtttcagagtcttcgtgccgccgtctacatctatctccattaggtctgaagatgactcttcacaccaacgacgttt
                                                                 IR1              IR1
       310       320       330       340       350       360       370       380       390       400
aaggtctctatcctactcctagcttgcaatacctggcttgcaatacctggagcatcgtgcacgatgattggatactgtggaggaggagtgtttgctgatt

       410       420       430       440       450       460       470       480       490       500
tagagctcccggttgggtgatttgacttcgatttcagtttaggcttgttgaaattttcaggttccattgtgaagcctttagagcttgagcttccttcca

       510       520       530       540       550       560       570       580       590       600
tgttaatgccttgatcgaattctcctagagaaagggaagtcgatctctgagtattgaaatcgaagtgcacatttttttcaacgtgtccaatcaatcca
                IR2                 IR2
       610       620       630       640       650       660       670       680       690       700
caaacaaagcagaagacaggtaatctttcatacttatactgacaagtaatagtcttaccgtcatgcataataacgtctcgttccttcaagaggggttttc

       710       720       730       740       750       760       770       780       790       800
cgacatccataacgacccgaagcctcatgaaagcattagggaagaacttttggttcttcttgtcatggcctttataggtgtcagccgagctcgccaattc

       810       820       830       840       850       860       870       880       890       900
ccgtccgactggctccgcaaaatattcgaacggcaagttatggacttgcaaccataactccacggtattgagcaggacctattgtgaagactcatctcat

       910       920       930       940       950       960       970       980       990      1000
ggagcttcagaatgtggttgtcagcaaaccaatgaccgaaatccatcacatgacggacgtccagtgggtgagcgaaacgaaacaggaagcgcctatcttt

      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
cagagtcgtgagctccacaccggattccggcaactacgtgttgggcaggcttcgccgtattagagatatgttgaggcaagacccatctgtgccactcgta

      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
caattacgagagttgtttttttgtgattttcctaagtttctcgttgatggtgagctcatattctacatcgtatggtctctcaacgtcgtttcctgtcat
```

FIGURE 4B

```
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
ctgatatcccgtcatttgcatccacgtgcgccgcctcccgtgccaagtccctaggtgtcatgcacgccaaattggtggtggtgcgggctgccctgtgctt
                       ABRE

     1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
cttaccgatgggtggaggttgagtttgggggtctccgcggcgatggtagtgggttgacggtttggtgtgggttgacggcattgatcaatttacttcttgc
                                                      R1                R1

     1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
ttcaaattctttggcagaaaacaattcattagattagaactggaaaccagagtgatgagacggattaagtcagattccaacagagttacatctcttaaga

     1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
aataatgtaacccctttagactttatatatttgcaattaaaaaaataatttaacttttagactttatatatagtttaataactaagtttaaccactcta
                   R2                                         R2

     1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
ttatttatatcgaaactatttgtatgtctcccctctaaataaacttggtattgtgtttacagaacctataatcaaataatcaatactcaactgaagtttg

     1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
tgcagttaattgaagggattaacggccaaaatgcactagtattatcaaccgaatagattcacactagatggccatttccatcaatatcatcgccgttctt

     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
cttctgtccacatatcccctctgaaacttgagagacacctgcacttcattgtccttattacgtgttacaaaatgaaacccatgcatccatgcaaactgaa
                                                                                 Legumin Vicilin

     1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
gaatggcgcaagaacccttcccctccatttcttatgtggcgaccatccatttcaccatctcccgctataaaacacccccatcacttcacctagaacatca
                                                       CAAT      TATA

     2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
tcactacttgcttatccatccaaaagatacccaccATGGCTAGATCATCAAGCCCTTTGCTTCTCTCACTCTGCATTTTCGCCATTCTCTTCCACTCTTC
                                    M  A  R  S  S  S  P  L  L  L  S  L  C  I  F  A  I  L  F  H  S  S
                                                Signal sequence

     2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
TCTGGGTAGGCAGCAATTCCAGCAGGGGAACGAGTGCCAGATCGACAGGATCGACGCATCCGAGCCGGACAAAACCATCCAGGCAGAAGCTGGCACCATC
 L  G  R  Q  Q  F  Q  Q  G  N  E  C  Q  I  D  R  I  D  A  S  E  P  D  K  T  I  Q  A  E  A  G  T  I

     2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
GAGGTATGGGACCAGAACCGCCAGCAATTCCAGTGCGCTGGTGTTGCCGTTGTAAGGCGCACCATTGAGCCCAAAGGTCTTCTCTTGCCTTTCTACAGCA
 E  V  W  D  Q  N  R  Q  Q  F  Q  C  A  G  V  A  V  V  R  R  T  I  E  P  K  G  L  L  L  P  F  Y  S
```

FIGURE 4C

```
       2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
ACACCCCTCAGCTCATCTACATCGTTCAAGgtataaattaaatcagttcatacaatgataaccaccacttcgaatgtatttatcaaatatcaatgatcga
N  T  P  Q  L  I  Y  I  V  Q

       2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
tgcacctgtatgtgttgtgtatattcagGTAGGGGAGTTACAGGAATCATGTTCCCAKGATGTCCAGAGACATTCGAGGAATCCCAGCAGCAAGGACAAC
                            G  R  G  V  T  G  I  M  F  P  X  C  P  E  T  F  E  E  S  Q  Q  Q  G  Q

       2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
AGGGCCAACAGGGTAGTTCCCAAGACCAGCACCAGAAGATCCGCCGCTTCCGTGAAGGTGACGTCATTGCCGTCCCTGCCGGTGTAGCCCACTGGTCCTA
 Q  G  Q  Q  G  S  S  Q  D  Q  H  Q  K  I  R  R  F  R  E  G  D  V  I  A  V  P  A  G  V  A  H  W  S  Y

       2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
CAACGATGGCAACGAACCAGTCATGGCCATTGTTGTCCATGACACTTCCAGCCACCTCAACCAACTGGACAACAACCCCAGGgtatataagcattgccgt
  N  D  G  N  E  P  V  M  A  I  V  V  H  D  T  S  S  H  L  N  Q  L  D  N  N  P  R

       2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
agttgctaataaattgcacacaattggaactctattttcagtatctaataactttttcctttttggcagAACTTCTACTTGGCAGGAAACCCGAGAGAC
                                                                      N  F  Y  L  A  G  N  P  R  D

       2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
GAGTTCGAACAATCGCAGCAAGGAGGCAGGCTGAGCCGTGGGGAGAGTGAAGGTGGACGAGGACGCAGGGAACCTCTTCAACCTGCAACAACCTCTTCTT
 E  F  E  Q  S  Q  Q  G  G  R  L  S  R  G  E  S  E  G  G  R  G  R  R  E  P  L  Q  P  A  T  T  S  S

       2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
GCGGAATCGACTCCAAGCTCATCGCGGAGGCGTTCAATGTCGACGAGAACGTGGCAAGGAGGCTACAGAGCGAGAACGACAACAGAGGCCAGATCGTCCG
C  G  I  D  S  K  L  I  A  E  A  F  N  V  D  E  N  V  A  R  R  L  Q  S  E  N  D  N  R  G  Q  I  V  R

       3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
AGTCGAAGGCGAGCTCGACATCGTCAGACCTCCGACCAGTATCCAGGAGGAGTCACAGGAGCAGGGAGGTCGTGGTGGTGGCCGCTACTACTCCAATGGA
  V  E  G  E  L  D  I  V  R  P  P  T  S  I  Q  E  E  S  Q  E  Q  G  G  R  G  G  G  R  Y  Y  S  N  G

       3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
GTGGAGGAGACCTTCTGCTCCATGAGACTAATTGAGAACATCGGCGATCCTTCTCGGGCAGACATTTTCACTCCAGAAGCCGGCCGCGTTAGATCCCTCA
 V  E  E  T  F  C  S  M  R  L  I  E  N  I  G  D  P  S  R  A  D  I  F  T  P  E  A  G  R  V  R  S  L
```

FIGURE 4D

```
       3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
ACAGCCACAACCTCCCCGTCCTGCAATGGATCCAGCTTAGCGCCGAGAGAGGCGTTCTCTACAATgtatagatctcactcacgcaccaactctaaattga
N  S  H  N  L  P  V  L  Q  W  I  Q  L  S  A  E  R  G  V  L  Y  N

       3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
atccctaattatttaattcaccgatatctgaccgaccggtttgaattttgtagGAAGCGATCAGGCTGCCGCACTGGAACATCAACGCACACAGCATAGT
                                                      E  A  I  R  L  P  H  W  N  I  N  A  H  S  I  V

       3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
GTACGCGATCAGAGGACAAGCCAGAGTCCAGATCGTGAACGAGGAAGGGAATTCGGTGTTCGATGGAGTGCTGCAGGAAGGACAGGTGGTGACGGTGCCG
 Y  A  I  R  G  Q  A  R  V  Q  I  V  N  E  E  G  N  S  V  F  D  G  V  L  Q  E  G  Q  V  V  T  V  P

       3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
CAGAACTTCGCGGTGGTAAAGAGATCCCAGAGCGAGAGGTTTGAGTGGGTGGCGTTCAAGACCAACGACAACGCGATGGTGAACTCGCTAGCCGGGAGGA
 Q  N  F  A  V  V  K  R  S  Q  S  E  R  F  E  W  V  A  F  K  T  N  D  N  A  M  V  N  S  L  A  G  R

       3610      3620      3630      3640      3650      3660      3670      3680      3690      3700
CATCGGCAGTAAGGGCGATCCCCGCGGATGTACTGGCTAACGCCTGGAGGGTGTCGCCGGAGGAGGCGAGGAGGGTGAAGTTCAACAGGCAGGAGACTCA
T  S  A  V  R  A  I  P  A  D  V  L  A  N  A  W  R  V  S  P  E  E  A  R  R  V  K  F  N  R  Q  E  T  H

       3710      3720      3730      3740      3750      3760      3770      3780      3790      3800
CTTGGCTAGCACCAGGGGCCAGTCCAGGTCGCCCGGGAGGTTGAATGTCGTCAAGGAGGTGATCAACTTGCTTATGTAAatgtgacggtgaaataataa
 L  A  S  T  R  G  Q  S  R  S  P  G  R  L  N  V  V  K  E  V  I  N  L  L  M  *

       3810      3820      3830      3840      3850      3860      3870      3880      3890      3900
cggtaaaatatatgtaataataataataaagccacaaagtgagaatgaggggaaggggaaatgtgtaatgagccagtagccggtggtgctaatttg

       3910      3920      3930      3940      3950      3960      3970      3980      3990      4000
tatcgtattgtcaataaatcatgaattttgtggtttttatgtgtttttttaaatcatgaattttaaattttataaaataatctccaatcggaagaacaac

       4010      4020      4030      4040      4050      4060      4070      4080      4090      4100
attccatatccatggatgtttctttacccaaatctagttcttgagaggatgaagcatcaccgaacagttctgcaactatccctcaaaagctttaaaatga

       4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
acaacaaggaacagagcaacgttccaaagatcccaaacgaaacatattatctatactaatactatattattaattactactgcccggaatcacaatccct
```

FIGURE 4E

```
        4210      4220      4230      4240      4250      4260      4270      4280      4290      4300
gaatgattcctattaactacaagccttgttggcggcggagaagtgatcggcgcggcgagaagcagcggactcggagacgaggccttggatgagcagagtc

        4310      4320      4330      4340      4350      4360      4370      4380      4390      4400
tttacctgccagggcgtgaaggggaagagcggccttctggagtaggagttcagcaagcggcggttccttggcggagtaagcggacgtaagggtggntgtc

        4410      4420      4430      4440      4450      4460      4470      4480      4490      4500
gacgtcntcgtttcnggaggcgnattcatgaagggttaaagtcanatctgtagctctcgagtgctcagggagccnaaagacgttgggaaaccgtcgncgt

        4510      4520      4530      4540      4550      4560      4570      4580      4590      4600
ttggggcatcagtcngcggggcacgcttccctcctgctgctccanaancnangtanatttaaaaganatgggaaattaantaatggnaatnannaggagg

        4610      4620      4630      4640      4650      4660      4670      4680      4690      4700
attgnaacggtcnganccgnangaanagttttttanngggtttaaatactgggggagtngnagccngccnctggttccngtgtagangaaaccaagnnccgg

        4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
gaggnttncannngnnagggagaaaaagganncatttnannangcngagggacatgaancggtacngagctgnggttcannnancggcgnnnggnagtcc

        4810      4820      4830      4840      4850      4860      4870      4880      4890      4900
cnngggaccnggntggggtnanaagggaanggaacattnggtngnangganaanaccnttttacnattgcctttgcaggnnngtntnggcncntncgggt

        4910      4920      4930      4940      4950      4960      4970      4980      4990
nacatnccgctgcatgggctttggggngccnanaggnagccncangggnanncngccnccttgtncangncgctnaagttcnattgtanatggncgttg
```

bonds, it is more stable than an A•T base pair, which has only two hydrogen bonds. The more G•C base-pairs are contained in a DNA, the greater the energy that is needed to separate the two strands; the T_m increases ~0.4°C for every 1% increase in G•C content. When DNA is in solution under approximately physiological conditions, the T_m usually lies in a range of 85–95°C. (A DNA that is 40% G•C—a value typical of mammalian genomes—denatures with a T_m of about 87°C under approximately physiological conditions; a DNA that is 60% G•C has a T_m of ~95°C under the same conditions.) Thus without intervention from cellular systems, duplex DNA is stable at the temperature prevailing in the cell.

Nucleic acids hybridize by base pairing

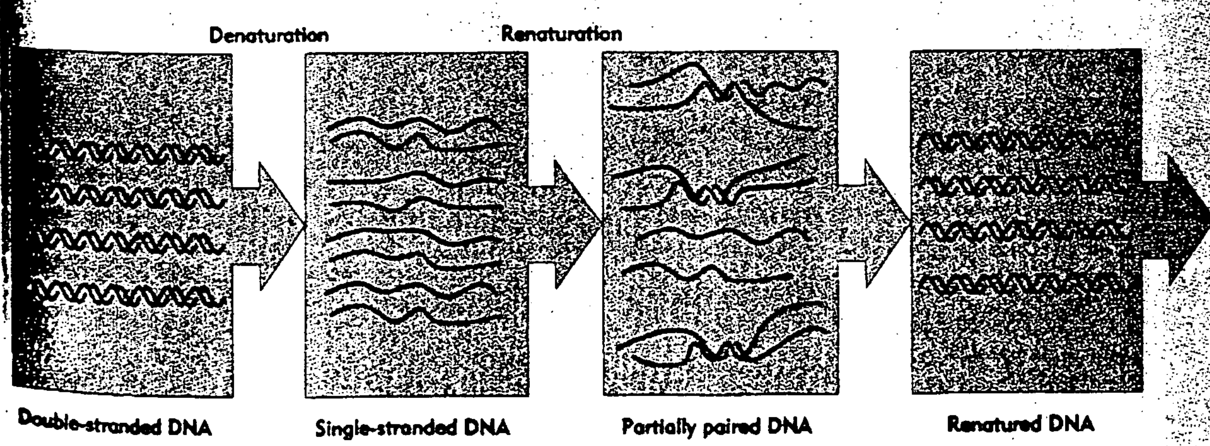
Nucleic acid sequences can be assessed in terms of either similarity or complementarity.

- *Similarity* between two sequences is given in principle by the proportion of bases (for single-stranded sequences) or base pairs (for double-stranded sequences) that is identical. Without determining the actual sequences, however, there is no direct way to measure similarity.

- *Complementarity* is determined by the rules for base pairing between A•T and G•C. In a perfect duplex of DNA, the strands are precisely complementary. If we compare two different but related double-stranded molecules, therefore, each strand of the first molecule will be similar to one strand of the second molecule and will be (partly) complementary to the other strand of the second molecule. Complementarity can be measured directly by the ability of two single-stranded nucleic acids to base pair with each other. If double-stranded molecules are denatured into single strands, the complementarity between the single strands can be used to indicate the similarity between the original duplex molecules.

Figure 5.2

Denatured single strands of DNA can renature to give the duplex form.



It is possible to measure complementarity because the denaturation of DNA is reversible under appropriate conditions. The ability of the two separated complementary strands to reform into a double helix is called **renaturation.** It is illustrated in **Figure 5.2**.

Renaturation depends on specific base pairing between the complementary strands. The reaction takes place in two stages. First, single strands of DNA in the solution encounter one another by chance; if their sequences are complementary, the two strands base pair to generate a short double-helical region. Then the region of base pairing extends along the molecule by a zipper-like effect to form a lengthy duplex molecule. Renaturation of the double helix restores the original properties that were lost when the DNA was denatured. Renaturation describes the reaction between two complementary sequences that were separated by denaturation. However, the technique can be extended to allow any two complementary nucleic acid sequences to **anneal** with each other to form a duplex structure.

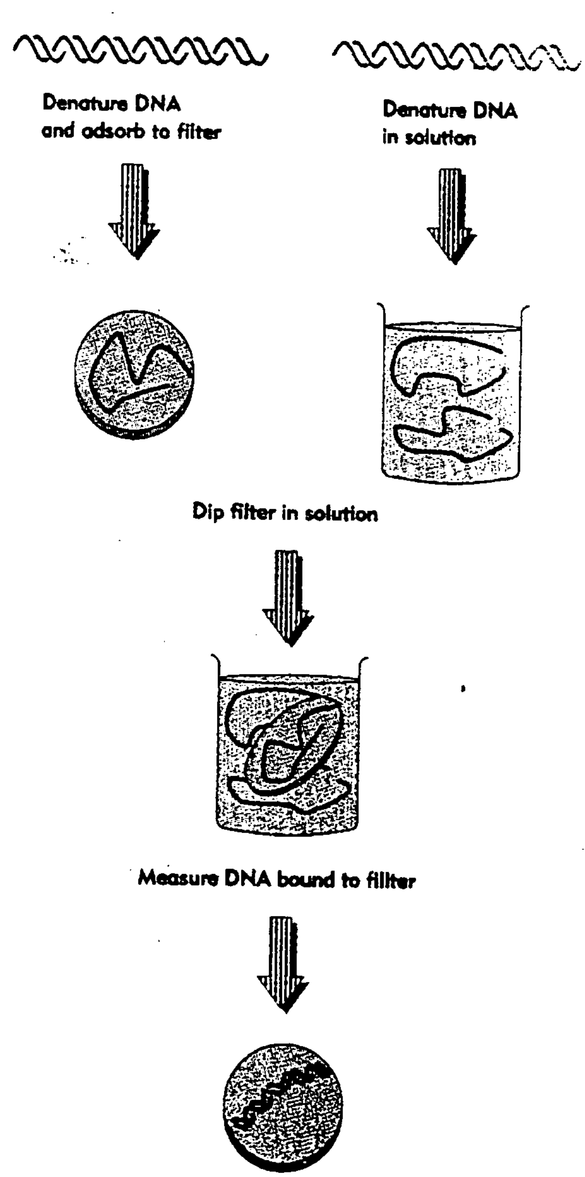
The reaction is generally described as **hybridization** when nucleic acids from different sources are involved, as in the case when one preparation consists of DNA and the other consists of RNA. *The ability of two nucleic acid preparations to hybridize constitutes a precise test for their complementarity since* only *complementary sequences can form a duplex structure.*

The principle of the hybridization reaction is to expose two single-stranded nucleic acid preparations to each other and then to measure the amount of double-stranded material that forms. There are two common ways of performing the reaction: **solution (liquid) hybridization and filter hybridization.**

Liquid hybridization is described by its name: the two preparations of single-stranded DNA are mixed together in solution. When large amounts of material are involved, the reaction can be followed by the change in optical density. With smaller amounts of material, one of the preparations may carry a radioactive label, whose entry into duplex form is followed by determining the amount of double-stranded DNA containing the label. Double-stranded DNA can be assayed either by using chromatography to separate duplex DNA from single strands or by degrading all the single strands that have not reacted and then

Figure 5.3

Filter hybridization establishes whether a solution of denatured DNA (or RNA) contains sequences complementary to the strands immobilized on th filter.



measuring the amount of material that remains.

Solution hybridization is not an appropriate technique for investigating the relationship of two preparations if one or both consist of duplex DNA. The problem is that if two duplex DNA preparations are denatured and then the single strands are mixed, two types of reaction occur. The *original* complementary single strands can renature. Or each single strand can hybridize with a complementary sequence in the *other* DNA. The competition between the two reactions makes it difficult to assess the extent of hybridization.

This difficulty can be overcome by immobilizing one of the DNA preparations so that it cannot renature. Nitrocellulose filters have the useful property of adsorbing single strands of DNA but not RNA; and once a filter has been used to adsorb DNA, it can be treated to prevent any further adsorption of single strands.

Figure 5.3 illustrates the resulting procedure in which a DNA preparation is denatured and the single strands are adsorbed to the filter. Then a second denatured DNA (or RNA) preparation is added. This material adsorbs to the filter only if it is able to base pair with the DNA that was originally adsorbed. The usual form of the experimental procedure is to add a radioactively labeled RNA or DNA preparation to the filter, allowing the extent of reaction to be measured as the amount of radioactive label retained by the filter.

The extent of hybridization between two single-stranded nucleic acids can be taken in principle to represent their degree of complementarity. Two sequences need not be *perfectly* complementary to hybridize; if they are closely related but not identical, an imperfect duplex is formed in which base pairing is interrupted at positions where the two single strands do not correspond.

Single-stranded nucleic acids may have secondary structure

The stability of the double helix results from the hydrogen bonding between the complementary A•T and G•C pairs and also from interactions between the bases as they are 'stacked' above each other along the axis of the helix. These forces can be used to predict the stability of a double helix between two complementary sequences. Because RNA is the predominant single-stranded nucleic acid, the formation of double-stranded regions from a single strand is usually analyzed in terms of RNA, but the technique is equally valid for single-stranded DNA.

The primary structure of RNA is the same as that of DNA: a polynucleotide chain with 5'-3' sugar-phosphate links. Considered as a single strand, the molecule follows a random path in space, but base pairing within it can fix the location of one region relative to another.

When a sequence of bases is followed by a complementary sequence nearby in the same molecule, the chain may fold back on itself to generate an antiparallel duplex structure, called a **hairpin**. It consists of a base-paired, double-helical region, the **stem**, with a **loop** of unpaired bases at one end. **Figure 5.4** shows an example. When the complementary sequences are relatively distant in the molecule, their juxtaposition to form a double-stranded region essentially creates a stem with a very long single-stranded loop.

Our ability to measure secondary structure is rather crude. The *overall* extent of base pairing is reflected in the biophysical properties of a molecule. However, this does not reveal which individual regions are involved. Single-stranded and double-stranded regions have different susceptibilities to some **nucleases** (enzymes that degrade nucleic acids), and this provides a test for analyzing the involvement of particular regions in base